

4 Solução Proposta

A metodologia proposta nesta dissertação pode ser dividida em quatro etapas complementares que estão representadas no diagrama de blocos na Figura 3. Cada uma delas está descrita através do detalhamento de suas características e dos objetivos de cada técnica aplicada ao sistema de inferência da qualidade proposto.

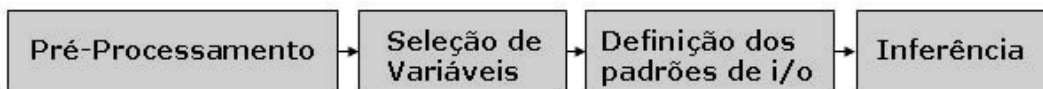


Figura 3: Diagrama de blocos do sistema de inferência da qualidade.

4.1. Pré-Processamento

Os dados da base histórica para implementar o método de inferência proposto são compostos por 12 arquivos digitais, organizados sob a forma de planilhas eletrônicas, uma para cada mês do ano em que os dados foram coletados. Em cada planilha, as colunas representam os instrumentos de medidas, são eles: 21 medidores de temperatura, 9 medidores de pressão, 10 medidores de vazão e 1 medidor de nível. Os valores de cada instrumento foram medidas de minuto em minuto, e armazenadas em cada linha da tabela. Além disso, uma última coluna foi inserida com dados anotados pela equipe de operação que registravam o valor de temperatura do ponto de ebulição de amostras de nafta craqueada, coletadas de 8 em 8 horas e analisadas em laboratórios especializados.

Na Tabela 1 a seguir, estão resumidos todos os instrumentos que foram utilizados na formação do banco de dados, incluindo identificações de campo, respectivas faixas de operação e unidades de engenharia.

Instrumento	TAG	Tipo	Faixa	Unidade
1	TIC39249	Temperatura	0 - 150	°C
2	TI39193	Temperatura	0 - 200	°C
3	TIC39195	Temperatura	0 - 300	°C
4	TI39196	Temperatura	0 - 200	°C
5	TIC39247	Temperatura	0 - 150	°C
6	TI39231	Temperatura	0 - 200	°C
7	TI39251	Temperatura	0 - 100	°C
8	TI39252	Temperatura	0 - 100	°C
9	TI39253	Temperatura	0 - 100	°C
10	TI39283	Temperatura	0 - 100	°C
11	TI39234	Temperatura	0 - 200	°C
12	TI39235	Temperatura	0 - 200	°C
13	TI39236	Temperatura	0 - 200	°C
14	TI39237	Temperatura	0 - 200	°C
15	TI39238	Temperatura	0 - 200	°C
16	TIC39239	Temperatura	0 - 200	°C
17	TI39232	Temperatura	0 - 300	°C
18	TI39233	Temperatura	0 - 100	°C
19	TI39245	Temperatura	0 - 100	°C
20	TI39279	Temperatura	0 - 100	°C
21	TI39246	Temperatura	0 - 150	°C
22	PI39205	Pressão	0 - 4	Kgf/cm ²
23	PI39206	Pressão	0 - 10	Kgf/cm ²
24	PI39207	Pressão	0 - 10	Kgf/cm ²
25	PI39209	Pressão	0 - 4	Kgf/cm ²
26	PI39228	Pressão	0 - 40	Kgf/cm ²
27	PI39277	Pressão	0 - 10	Kgf/cm ²
28	PI39201	Pressão	0 - 16	Kgf/cm ²
29	PI39200	Pressão	0 - 40	Kgf/cm ²
30	PI39224	Pressão	0 - 16	Kgf/cm ²
31	FIC39179	Vazão	0 - 8000	m ³ /d
32	FIC39180	Vazão	0 - 5000	m ³ /d
33	FI39230	Vazão	0 - 2600	m ³ /d
34	FIC39227	Vazão	0 - 1500	m ³ /d
35	FIC39173	Vazão	0 - 50000	m ³ /d
36	FIC39175	Vazão	0 - 30000	m ³ /d
37	FIC39174	Vazão	0 - 15000	m ³ /d
38	FFIC39223	Vazão	0 - 1500	m ³ /d
39	FI39188	Vazão	0 - 1500	m ³ /d
40	FIC39178	Vazão	0 - 8000	m ³ /d
41	LIC39155	Nível	0 - 100	%
42	U39PFE/NC	Análise Laboratorial	40 - 220	°C

Tabela 1: Escala e unidade dos instrumentos do banco de dados original.

A etapa de pré-processamento é fundamental para a estruturação desta metodologia, visto que, é nela que se eliminam os ruídos de processo e desvios de operação que influenciam diretamente no resultado do treinamento da rede neural e, posteriormente, no resultado final do processo. Entretanto, foi necessário subdividir as etapas de pré-processamento, conforme ilustrado na

Figura 4, para um melhor detalhamento das técnicas de correção implementadas.

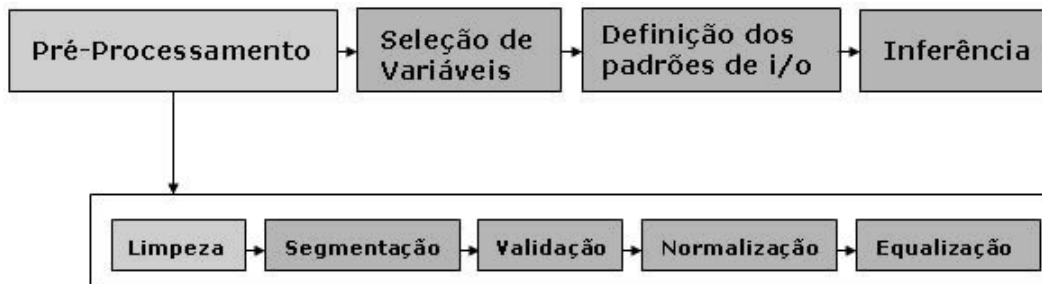


Figura 4: Etapas de pré-processamento.

4.1.1. Limpeza

Observando os dados coletados e avaliando as suas consistências lógicas, notou-se que era necessário fazer a remoção de ruídos e de informações redundantes. Este processo deveria idealmente envolver a subjetividade do conhecimento dos operadores da planta de produção, pois a partir da observação dos dados e dos gráficos de processo, eles poderiam propor correções significativas ao sistema. Contudo, como tal participação envolveria aspectos logísticos pouco viáveis, foram criadas rotinas para automatizar esta verificação. Com base na análise crítica do banco de dados, os valores de leitura foram avaliados, destacando aqueles considerados discrepantes à funcionalidade normal da planta de nafta.

Dentre os problemas encontrados nas planilhas de dados, foi possível destacar dois que representam a maioria deles, como: a incoerência nos valores de leitura, que se apresentavam zerados ou inalterados em um grande espaço de tempo, e ainda, valores instantâneos com resultados de 5 a 10 vezes maiores que o fundo de escala do respectivo instrumento.

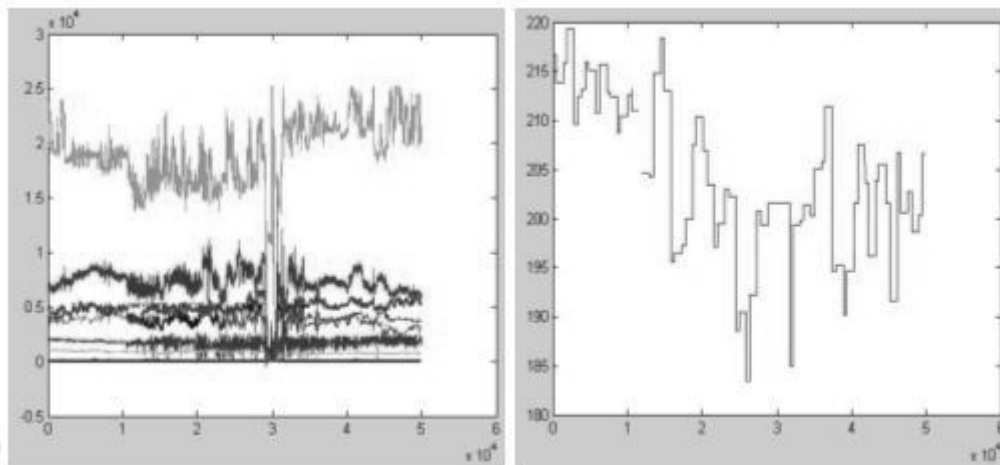
Para solucionar esta primeira etapa de limpeza dos dados, foi analisada a relação entre os valores de leitura de cada instrumento e sua respectiva escala. Alguns instrumentos mostraram um problema de mau dimensionamento, onde, por exemplo, existiam medidas variando na faixa de 0 a 100°C, porém utilizando equipamentos com faixas de operação de 0 a 1000 °C. Em termos práticos isto afeta a qualidade da medida, pois diminui a resolução, principalmente para pequenas variações. Sendo assim, os instrumentos TI39232, PI39200, FIC39227, FFIC39223 e FIC39173, que estavam com estas características,

foram retirados do banco de dados, para permitir uma análise mais acurada e segura dos resultados da metodologia.

Com a eliminação de cinco instrumentos, os dados de leitura dos 36 medidores restantes foram reavaliados, através de um algoritmo de inspeção que identificou cada instante cuja leitura estava fora da faixa de operação, sem valores descritos ou ainda com termos zerados. A marcação destes pontos foi efetuada substituindo as leituras danificadas pelo código NaN (Not a Number).

O segundo algoritmo de limpeza de dados possui um princípio de funcionamento simples. Primeiro, as informações de cada tabela mensal foram aglomeradas em uma matriz de dados única anual, onde as linhas representam as leituras de cada instante armazenadas por 12 meses, e as colunas os diversos instrumentos. O segundo passo consiste em encontrar os instantes com código NaN e substituí-los pelo respectivo valor de leitura do instante anterior. Desta forma, os sinais mantêm a coerência adequada para representar a realidade de leitura dos instrumentos de medição de campo.

O resultado desta primeira etapa de pré-processamento pode ser observado na Figura 5. As leituras ao longo de um ano dos 36 instrumentos representam as entradas do sistema, ao passo que os dados laboratoriais (última coluna de cada uma das 12 planilhas) representam a saída padrão que será utilizada, posteriormente, no treinamento da rede neural.



(a)

(b)

Figura 5: Dados de entrada (a) e de saída (b) após a limpeza de dados.

4.1.2. Segmentação

A necessidade de segmentar o banco de dados está relacionada à diferença de frequência existente entre a captura da leitura dos instrumentos (minuto em minuto) e o tempo de registro das análises de laboratório (8 em 8 horas). Em termos práticos, o sistema estaria considerando que a saída, isto é, o ponto de ebulição da nafta craqueada, é constante por períodos de 8 horas. Todavia, não necessariamente isto é garantido, uma vez que neste período podem ocorrer oscilações nas entradas.

Sendo assim, da mesma forma que foi descrito em [HALL02], para cada instante de transição da saída, foi definida uma janela de entrada englobando as leituras dos 36 instrumentos durante um intervalo de tempo anterior a este instante.

O algoritmo elaborado para compatibilizar as frequências entre os sinais de entrada e saída, ou seja, o algoritmo de segmentação, é iniciado com a análise da variável de saída Y através da sua derivada Yd . Durante a derivação, os intervalos cuja saída é constante são zerados, e os demais, que representam alguma variação linear, retornam valores constantes. Desta forma, os instantes que sofrem alguma variação são identificados e um gráfico contendo somente períodos de transição é gerado, passando a representar a saída Y , conforme pode ser observado na Figura 6.

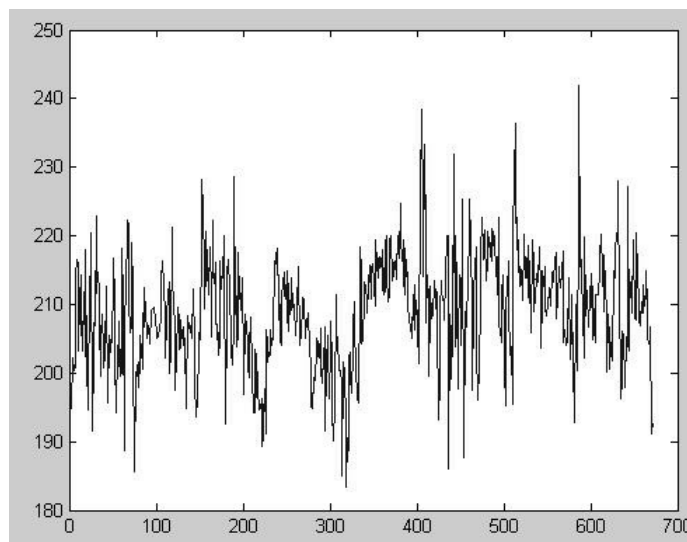


Figura 6: Amostra do sinal de saída depois da segmentação.

Uma vez segmentados os dados da nova matriz Y de saída, o próximo passo é executar o mecanismo para compatibilizar a matriz X dos dados de entrada. Isto foi possível, alocando, no instante de transição de cada ponto de saída, uma janela de entrada formada pelo intervalo dos 360 minutos que antecedem o respectivo instante de saída. O tamanho desta janela foi determinado heurísticamente.

No final da execução do algoritmo, o sistema encontrou 771 pontos de transição de saída e a mesma quantidade de janelas de entrada, que, por sua vez, foram armazenados em um vetor de saída e em matrizes com dimensões de 360 linhas por 36 colunas, respectivamente. A Figura 7 **Erro! A origem da referência não foi encontrada.** apresenta um exemplo simplificado da segmentação envolvendo somente dois dos instrumentos de campo (1 e 2).

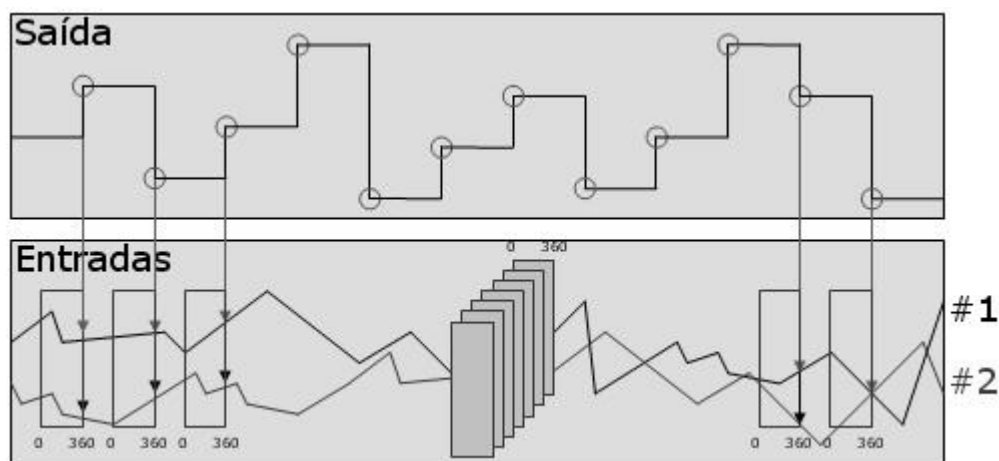


Figura 7: Processo de segmentação.

4.1.3. Validação

Após a segmentação dos dados de saída e da formação de suas respectivas janelas de entrada, foi observado que para a maioria dos dados de entrada de cada instrumento envolvido existe uma faixa natural de operação. Sendo assim, para melhorar a eficiência do sistema e reduzir sinais de pouca relevância, foi criado um procedimento para validar o banco de dados já segmentado. Isto se deu pela análise de janelas de entrada, cujo percentual de valores fora da faixa natural de operação excede um determinado limiar. O algoritmo que executa esta limpeza de dados funciona da seguinte maneira:

1. Uma janela é escolhida para ser analisada;

2. Calcula-se a média e o desvio padrão dos valores de leitura, de cada instrumento, correspondentes à janela escolhida. Supõe-se que os dados têm distribuição gaussiana centrada na média, como ilustrado na Figura 8. A partir dessa distribuição, é considerado que 95% dos dados são suficientes para representar o comportamento de cada instrumento. Desta forma, os limites máximo e mínimo de operação normal são definidos pela soma da média com 1,96 vezes o desvio padrão, e pela diferença entre a média e 1,96 vezes o desvio padrão, respectivamente.

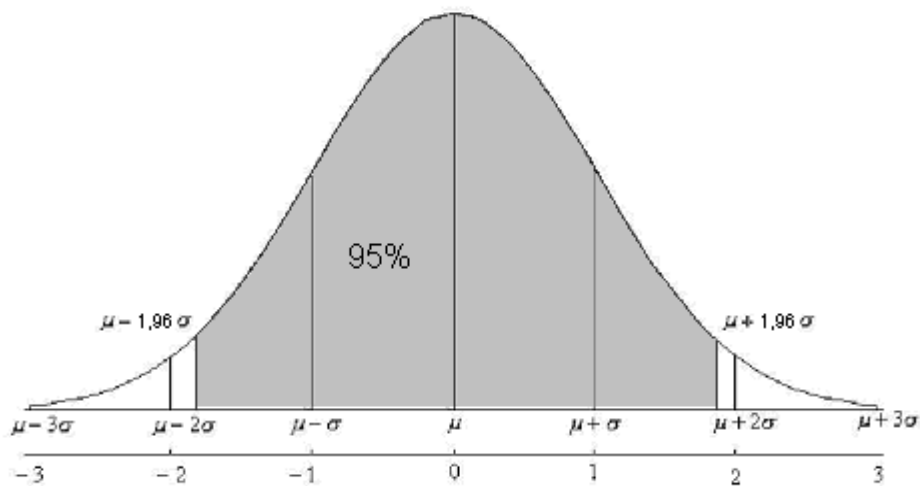


Figura 8: Densidade de probabilidade = Distribuição Gaussiana.

3. Com os limites definidos por instrumento e por janela, inicia-se o processo de validação, que marca os instantes cujos valores excedem a faixa limite, conforme é exemplificado na Figura 9.

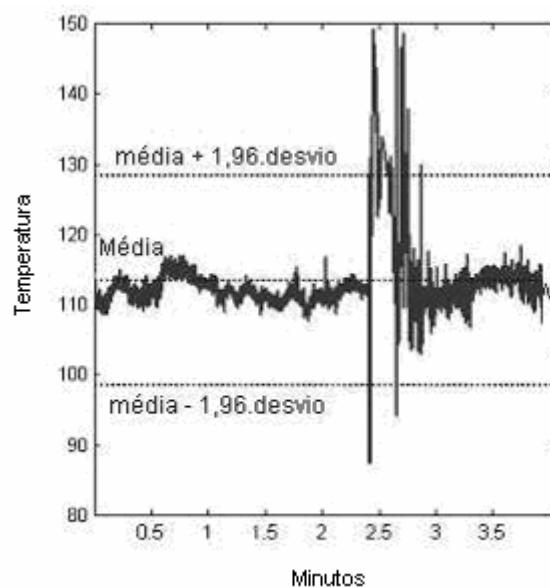


Figura 9: Exemplo de valores que excedem a faixa limite de um instrumento.

4. Após examinar e marcar todos os instantes fora da faixa limite, dois critérios foram criados para avaliar a importância de uma determinada janela. O primeiro avalia a quantidade de valores de leitura, de um determinado instrumento, que aparecem fora da faixa limite definida. Se esta quantidade exceder um percentual de 11% do total de instantes da janela, o instrumento será considerado fora do padrão de normalidade.
5. O segundo critério procura avaliar a janela integralmente, verificando se o número de instrumentos fora do padrão de normalidade, contidos em uma mesma janela, excede o limiar de 12% do total de instrumentos. Desta forma, as janelas que contêm um determinado nível de ruído, e não são adequadas para o processo de inferência, são identificadas e eliminadas.

Com base na avaliação destes critérios, o número de janelas iniciais, que era de 771, foi reduzido para 664. No capítulo 5 serão apresentados resultados considerando, ora um grupo de testes com todas as janelas de entrada, ora outro grupo com a quantidade de janelas reduzida.

4.1.4. Normalização

Como os dados adquiridos por vários sensores diferem muito em unidades e faixas de operação, é necessário empregar algum tipo de normalização antes de executar o treinamento da rede neural. Desta forma, o valor de leitura de um instrumento não estará sendo mais significativo que outro quando sua escala for diferente, ou seja, um medidor que trabalha de 0 a 150 não afetará mais o sistema que um outro com escala de 0 a 100.

A normalização mais usual consiste em subtrair o valor médio dos valores de cada sinal, e dividir este resultado pelo respectivo desvio padrão, obtendo assim um sinal normalizado com média zero e variância unitária. Contudo, tal método é de difícil implementação em tempo real, uma vez que se deseja estimar a média e a variância dos sinais a cada instante de forma incremental.

Desta forma, utilizou-se a faixa de variação nominal de cada sensor i para identificar o limite mínimo m_i e o limite máximo M_i de operação, e aplicar a normalização indicada pela Equação 5, onde x_i representa o valor a ser normalizado. Desta forma, são obtidos novos valores de leitura x_i^n para cada

sensor, variando sempre em uma faixa de zero a um, independentemente da escala do instrumento.

$$x_i^n = \frac{x_i - m_i}{M_i - m_i} \quad (5)$$

O resultado desta normalização pode ser visto nas Figuras 10 e 11 a seguir, que exemplificam os sinais correspondentes a um instrumento envolvido no processo, antes e depois do procedimento de normalização.

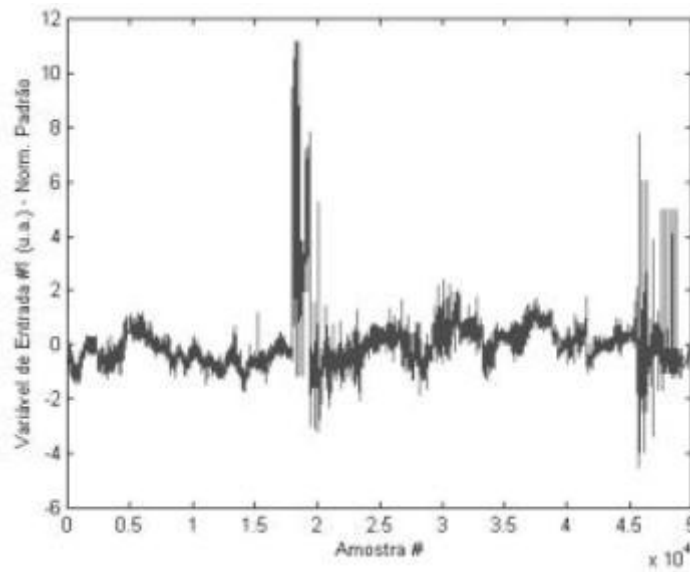


Figura 10: Sinal do instrumento # 1 antes da normalização.

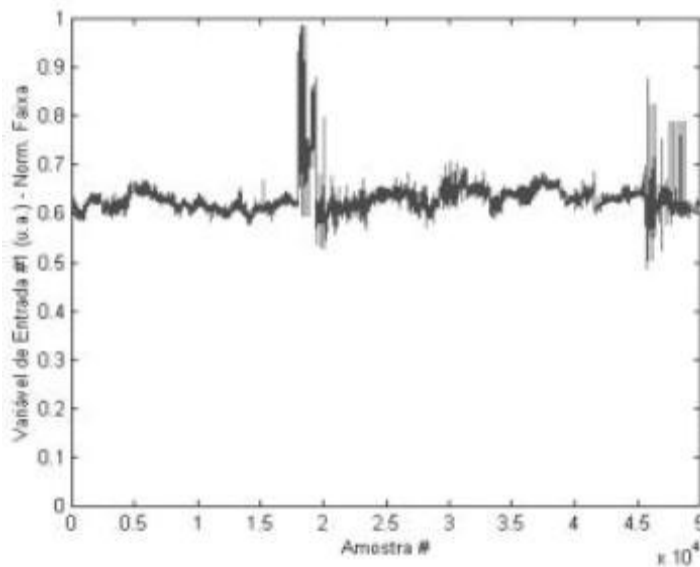


Figura 11: Sinal do instrumento # 1 depois da normalização.

4.1.5. Equalização

A equalização de dados, utilizando o histograma dos valores armazenados, é uma técnica muito aplicada em processamento de imagem, entretanto, pode servir também para definir uma normalização por partes, para os dados a serem inferidos no sistema proposto [HALL02]. Em termos práticos, esta ação melhora a faixa dinâmica de utilização do neurônio de saída, permitindo que a rede neural passe a ter uma melhor distribuição dos valores no aprendizado, evitando que os mesmos atuem somente em uma pequena porção da faixa ativa.

A fórmula que otimiza a faixa de variação da saída é expressa pela Equação 6, onde y_e representa a informação equalizada, y_n os dados já normalizados, L_{inf} o limite inferior e L_{sup} o limite superior que melhor representam a distribuição dos dados. Tais limites foram definidos pelos operadores da planta de produção de nafta, com base na experiência do processo, sendo $L_{inf} = 0,5$ e $L_{sup} = 0,9$. O resultado deste processamento pode ser visualizado na Figura 12, que exemplifica a distribuição da leitura da variável de saída do sistema.

$$y_e = \frac{y_n - L_{inf}}{L_{sup} - L_{inf}} \quad (6)$$

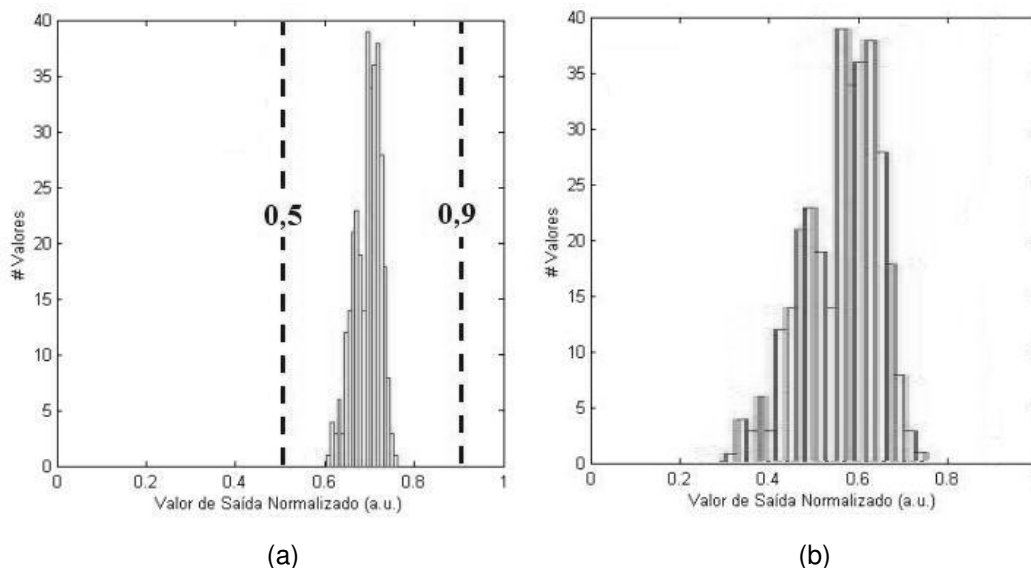


Figura 12: Histograma do sinal de saída antes (a) e depois (b) da equalização.

Após a conclusão das etapas de pré-processamento do banco de dados é possível preparar as entradas e saídas deste sistema para a realização do treinamento da rede neural. Contudo, para que a estrutura da rede fosse

otimizada, foi necessário reduzir o número de entradas deste sistema, inicialmente composto por 36 variáveis. Desta forma, as interações poderiam executar um menor número de cálculos e, conseqüentemente, consumir um tempo de processamento mais curto. Sendo assim, diversas técnicas foram estudadas, selecionadas e implementadas para melhorar o desempenho do sistema.

4.2. Seleção de Variáveis

As técnicas de seleção de variáveis podem definir, dentro de um conjunto de dados, aqueles que têm maior influência sobre a variável de saída inferida. Desta forma, o número de entradas pode ser reduzido sem que a relação com a saída sofra perda de informação e afete o processo de treinamento da rede neural.

O banco de dados armazenado inicialmente é composto por 664 matrizes de entradas, com dimensões de 360 linhas (instantes) e 36 colunas (instrumentos), e 664 saídas, que formam um vetor e são consideradas como padrões para o aprendizado do sistema de inferência.

Avaliando este conjunto de dados e prevendo a existência de instrumentos com informações redundantes, técnicas de seleção de variáveis, capazes de identificar as entradas que apresentam maior influência sobre a variável de saída, foram utilizadas, conforme ilustrado na Figura 13. A grande vantagem da aplicação destas técnicas está na redução da complexidade computacional e, conseqüentemente, no tempo de processamento do aprendizado da rede neural envolvida, além de melhorar a qualidade da inferência propriamente dita.

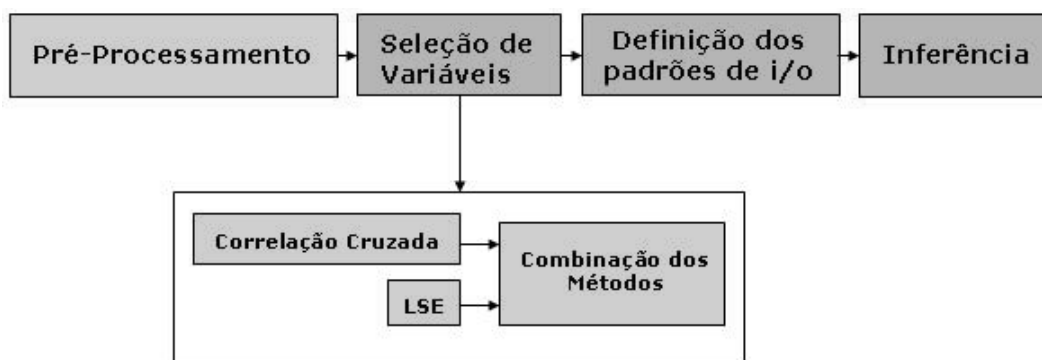


Figura 13: Etapas de seleção de variáveis.

Este processo seletivo poderia ser executado somente através do conhecimento operacional dos especialistas. Entretanto, informações empíricas do processo poderiam apontar para a escolha de atributos cujo impacto sobre a inferência não é conhecido. Sendo assim, foi proposto o uso de métodos matemáticos, que garantem maior segurança na seleção das variáveis de entrada em relação à saída inferida.

A seguir estão apresentados métodos de seleção de variáveis independentes de modelo, tais como a correlação cruzada e o LSE (Least Squares Estimator). O método PCA ([SLAMA88], [CONTRERAS02]), muito utilizada na área petroquímica desconsidera o relacionamento da matriz de entrada com a matriz de saída, limitando-se apenas à análise das entradas entre si. Apesar de ter sido implementado, mostrou resultados pouco promissores, não apresentando vantagens para o caso estudado, sendo, portanto, descartado.

4.2.1. Correlação Cruzada

Esta técnica aponta quais variáveis de entrada estão melhor relacionadas com uma determinada variável de saída. No caso deste trabalho, foi necessário o uso de um algoritmo de correlação cruzada adaptado, que considera as particularidades dos sinais com diferentes freqüências de amostragem.

Com base no processo implantado em [HALL02], foi utilizada a base de dados validada na etapa anterior e calculados os valores de correlação de cada variável de entrada e saída envolvidas no sistema. O algoritmo que implementa esta correlação cruzada modificada se baseia na Equação 7 a seguir.

$$c_j(\tau) = \frac{\sum_{i=1}^{N_y} (z_\tau[i] - \bar{z}_\tau) * (y[i] - \bar{y})}{N_y \cdot \partial z_\tau \cdot \partial y}, \quad (7)$$

$c_j(\tau)$ é o resultado da correlação cruzada calculada para o instrumento j , com atraso τ , que, por sua vez, representa cada instante da janela de entrada analisada; $z_\tau[i]$ representa os i -ésimo valor de entrada de um determinado instrumento com atraso τ ; $y[i]$ é o i -ésimo valor da própria saída; N_y é o

número de dados de saída e $\sigma_{z_\tau} \cdot \sigma_y$ são os desvios padrão das entradas e saídas respectivamente.

O procedimento calcula a correlação das 36 variáveis de entrada presentes no sistema com a variável de saída, para cada instante da janela analisada (atraso), de 1 a 360. Por exemplo, considerando a análise da variável 1 para o atraso 1, calcula-se a média e o desvio dos valores de entrada e de saída no instante 1 para as 664 janelas. De posse destes dados, a fórmula de correlação modificada calcula o somatório dos produtos entre as diferenças do valor correspondente ao instante 1 e sua média, e do valor de saída e sua média. Este somatório é dividido pelo produto dos desvios padrão de entrada e saída pelo número de janelas envolvidas no processo; temos, então, o resultado da correlação para o atraso 1.

O processo se repete até que os 360 pontos da janela sejam verificados. Desta forma, será formada uma curva de correlação cruzada com os dados da variável 1, e assim será definida a correlação máxima da variável, determinando inclusive o atraso, correspondente a esta variável, conforme pode ser visto na Figura 14 a seguir.

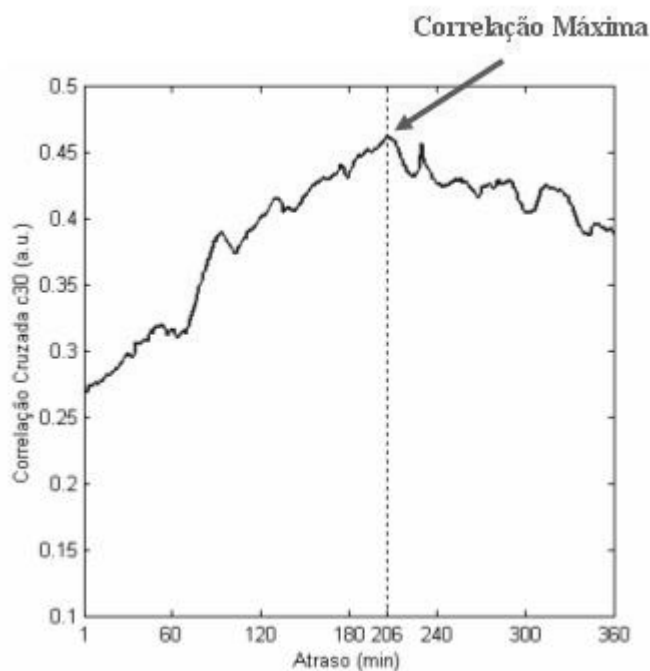


Figura 14: Correlação cruzada máxima de uma janela.

Após o processamento das 36 variáveis, o resultado foi computado nos dois gráficos apresentados na Figura 15, onde é possível observar os valores da correlação máxima para cada instrumento e seu respectivo atraso. Esta

informação será posteriormente utilizada para combinar resultados com outra técnica de seleção e preparar os dados para treinamento da rede neural de inferência.

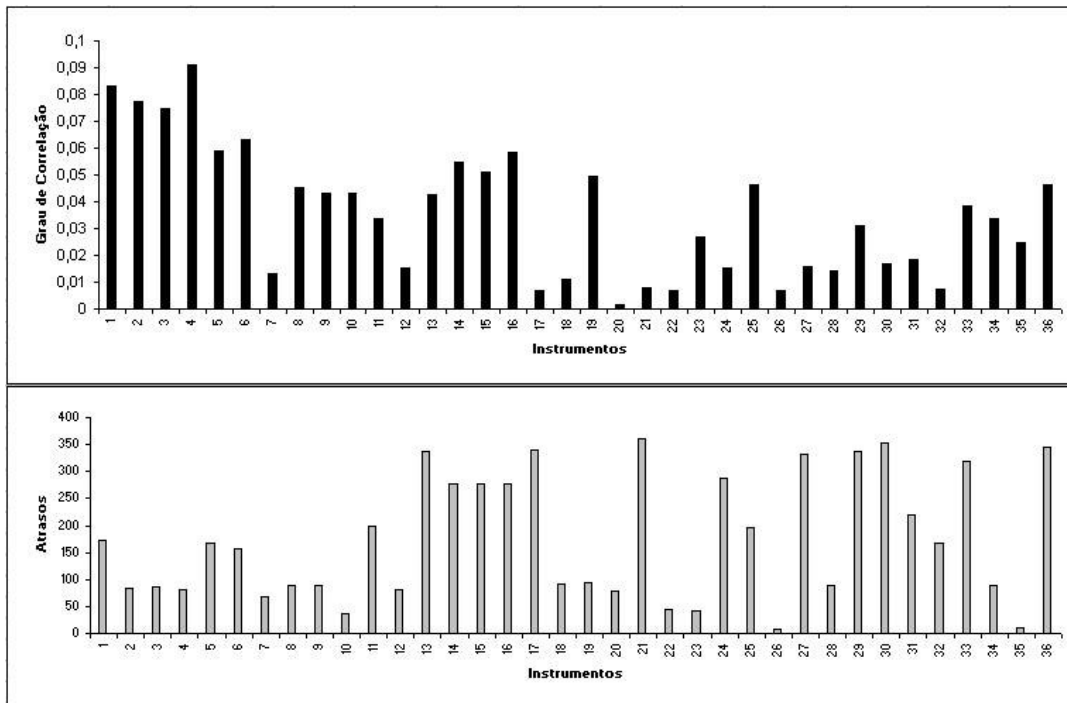


Figura 15: Valores da correlação cruzada máxima de cada instrumento e seus respectivos atrasos.

4.2.2. LSE (Least Squares Estimator)

O estimador por mínimos quadrados (LSE) é um método que busca expressar o comportamento das variações da saída y , em função das variações das diversas entradas X , por meio da relação $\Delta y = \Delta X \cdot b$, que será abordada mais adiante. Desta forma, é possível linearizar relações não-lineares por meio de uma regressão multivariada, cujos coeficientes são determinados por meio de otimização por mínimos quadrados. Onde, e representa o erro residual, b o vetor de coeficientes de regressão, ambos relativos à janela n no instante i , e c o número total de instantes em cada janela, neste caso, 360.

Na prática, o algoritmo LSE ([CONTRERAS02], [BAFFI00]) pode ser executado em 5 etapas consecutivas para cada atraso i , conforme descrito adiante.

1. Cria-se uma matriz X_i [664 x 36] com as informações das janelas de entrada relativas ao instante i . Nesta matriz, cada coluna representa um determinado instrumento de 1 a 36, e cada linha, o valor de leitura do instante i deste instrumento, para cada uma das 664 janelas.
2. Uma vez que o vetor de saída y apresenta um valor para cada uma das 664 janelas de entrada, considere y_n o valor relativo à janela n . Calcule-se, então, o vetor Δy das variações dos valores de saída, através da expressão $\Delta y_n = y_{n+1} - y_n$, onde n varia de 1 a 663.
3. De forma análoga, calcula-se a matriz ΔX^i , para cada instante i , através da expressão $\Delta X^i_{n,m} = X^i_{n+1,m} - X^i_{n,m}$, onde n varia de 1 a 663 e m representa cada um dos 36 instrumentos.
4. De posse de Δy e ΔX^i , calcula-se o vetor de coeficientes de regressão correspondente ao atraso i , b_i , através da expressão $\Delta y = \Delta X^i \cdot b_i$. Contudo, para isto, é necessário encontrar a matriz inversa de ΔX^i que não existe, visto que a mesma não é quadrada. Sendo assim, calcula-se a pseudo-inversa de ΔX^i , obtendo a expressão $b_i = (\Delta X^{i^T} \cdot \Delta X^i)^{-1} \cdot \Delta X^{i^T} \cdot \Delta Y$. Observe que b_i será um vetor com 36 linhas, ou seja, para cada instante i tem-se um vetor com os valores dos coeficientes de regressão de cada instrumento.
5. Calculados os vetores de coeficientes b_i , o resultado final gera um gráfico para cada instrumento semelhante ao apresentado na Figura 14. Este gráfico é construído a partir do coeficiente correspondente àquele instrumento obtido por cada atraso (variando entre 0 e 360). Da mesma forma que na correlação cruzada, foi utilizado o valor máximo de cada gráfico para representar o grau de relevância do LSE para cada instrumento, conforme pode ser visto na Figura 16.

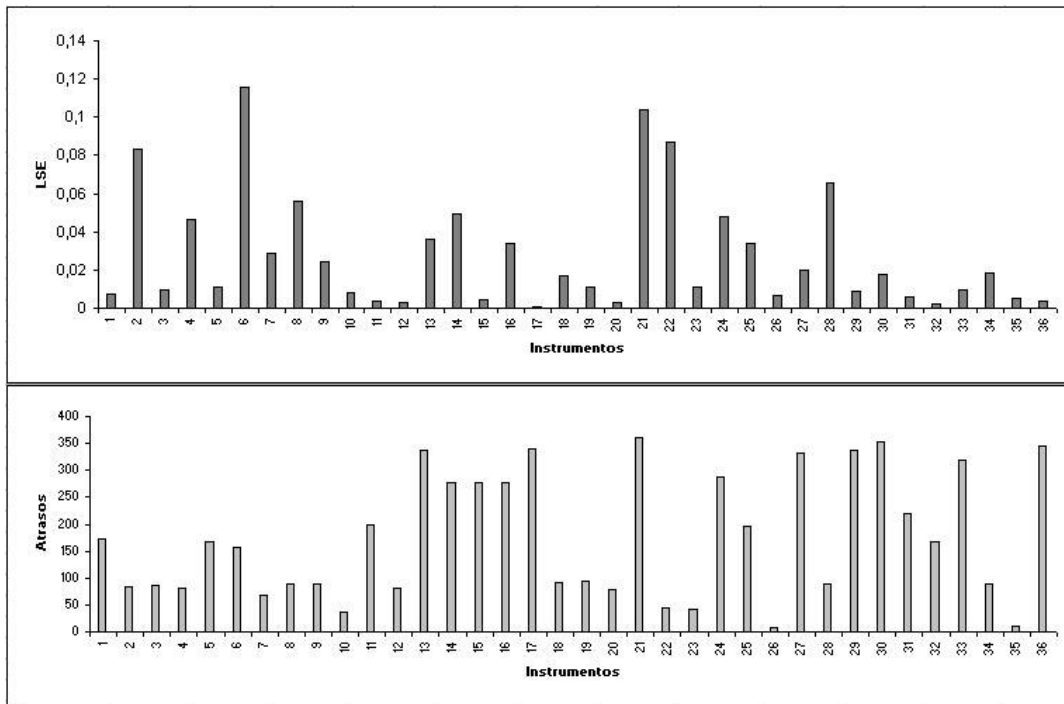


Figura 16: Valores do LSE de cada instrumento e seus respectivos atrasos.

4.2.3. Combinação dos Métodos

Buscando uma forma coerente de representar a união dos dois métodos de seleção apresentados anteriormente, foi elaborado um algoritmo que reúne para cada instrumento avaliado, através da soma dos valores absolutos, os graus de relevância encontrados pelas duas técnicas de seleção.

Todavia, para compatibilizar os graus de relevância, evitando que um método tenha mais influência que o outro, os resultados foram normalizados segundo seus valores absolutos, de forma que representassem uma distribuição normal. O grau normalizado de cada variável é obtido pela razão entre o grau atual de um dado instrumento e o somatório dos graus de todos os demais.

Concluída esta primeira fase, o próximo passo foi executar uma heurística para avaliar quantas e quais variáveis seriam suficientes para representar as informações relevantes ao processo. Sendo assim, foi estabelecido que todas as variáveis com grau de relevância inferior a 4,5% do somatório dos demais termos seriam eliminadas, e as exceções seriam, então, as variáveis selecionadas.

Na Figura 17, pode ser observada a combinação dos resultados, em valor absoluto, de ambas as técnicas de seleção de variáveis aplicadas, com seus respectivos atrasos.

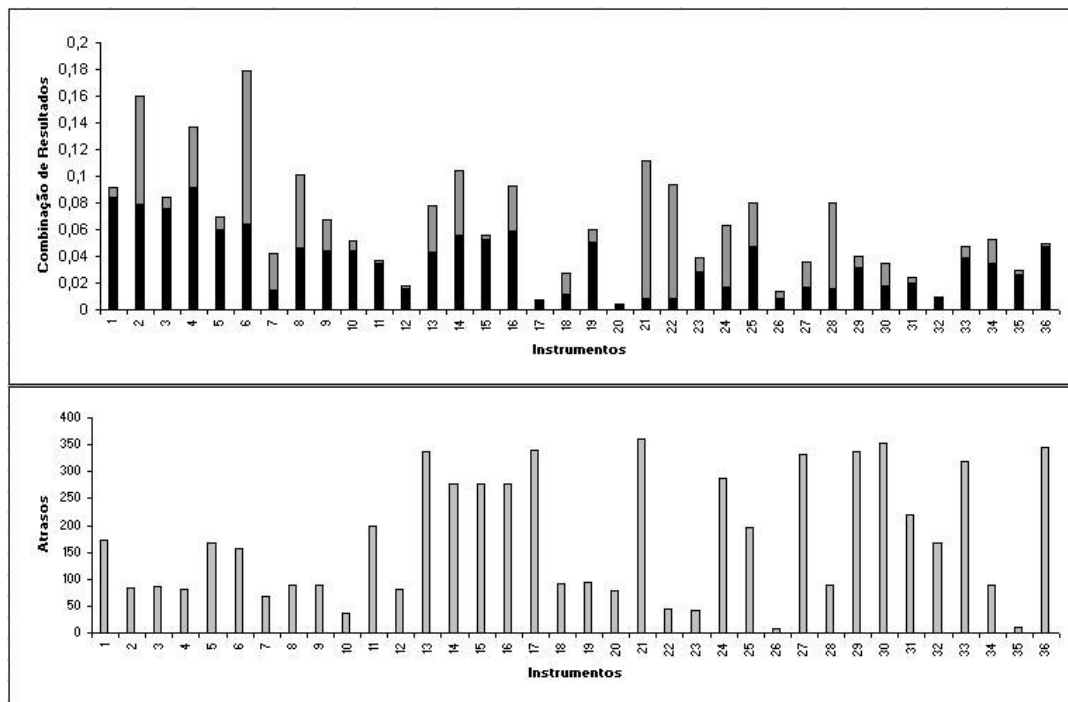


Figura 17: Combinação dos resultados das duas técnicas de seleção de variáveis.

A Tabela 2 apresenta os graus de relevância de cada método de seleção, assim como os resultados combinados. Além disto, destaca a ordem de relevância de cada instrumento, que foi encontrada a partir da ordenação decrescente dos valores obtidos na combinação dos métodos.

Como resultado, foram selecionadas cinco variáveis do banco de dados: 2, 4, 6, 8 e 14, que representam medidores de temperatura com as seguintes referências TIC39193, TI39196, TI39231, TI39252 e TI39237, sendo assim, as 664 janelas de entrada passam a ter dimensão [360x5].

Vale salientar que nesta etapa foram feitos testes e treinamentos para avaliar a real influência destas variáveis sobre o aprendizado de uma rede neural. Foi verificado que, ao se retirar qualquer uma destas cinco variáveis do processo, o treinamento da rede neural ficava comprometido. Em contrapartida, se qualquer uma das 31 variáveis restantes fossem retiradas, o resultado final dos testes seria pouco afetado. Isto comprovou que a técnica de seleção de variáveis foi bem sucedida.

Instrumento	Correlação	LSE	Combinação	Ordem de Relevância
1	0,0834	0,0077	0,0911	9°
2	0,0774	0,0831	0,1605	2°
3	0,0746	0,0098	0,0844	10°
4	0,0908	0,0462	0,137	3°
5	0,0587	0,0113	0,07	14°
6	0,0632	0,116	0,1792	1°
7	0,0132	0,0286	0,0418	23°
8	0,0452	0,0562	0,1014	6°
9	0,0432	0,0244	0,0676	15°
10	0,043	0,0082	0,0512	20°
11	0,0336	0,0034	0,037	26°
12	0,0151	0,0032	0,0183	32°
13	0,0424	0,0358	0,0782	13°
14	0,0548	0,0493	0,1041	5°
15	0,0511	0,0047	0,0558	18°
16	0,0584	0,0341	0,0925	8°
17	0,0068	0,001	0,0078	35°
18	0,011	0,0167	0,0277	30°
19	0,0493	0,0111	0,0604	17°
20	0,0018	0,0028	0,0046	36°
21	0,0078	0,1039	0,1117	4°
22	0,007	0,0872	0,0942	7°
23	0,0271	0,0114	0,0385	25°
24	0,0155	0,0478	0,0633	16°
25	0,0464	0,034	0,0804	11°
26	0,007	0,0064	0,0134	33°
27	0,0158	0,0198	0,0356	27°
28	0,0144	0,0657	0,0801	12°
29	0,031	0,0085	0,0395	24°
30	0,0169	0,0176	0,0345	28°
31	0,0185	0,0056	0,0241	31°
32	0,0076	0,0022	0,0098	34°
33	0,0382	0,0093	0,0475	22°
34	0,0337	0,0186	0,0523	19°
35	0,0249	0,0049	0,0298	29°
36	0,0463	0,0034	0,0497	21°

Tabela 2: Resultado da seleção de variáveis.

4.3. Padrões de Treinamento

Uma vez selecionadas as variáveis com maior relevância para o sistema, é preciso executar uma etapa intermediária para preparar os dados de entrada e saída que serão utilizados no treinamento da rede neural, conforme pode ser visto na Figura 18.

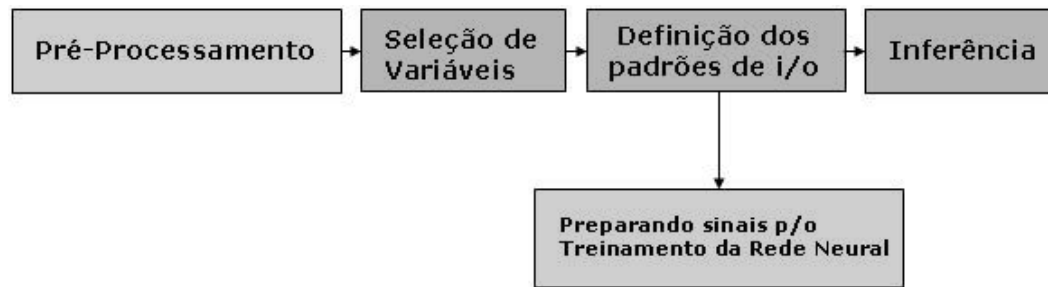


Figura 18: Definição dos padrões de entrada e saída.

A rede neural utilizada é do tipo multilayer perceptron com treinamento por backpropagation [HAYKIN94] e para o seu treinamento é necessária uma única matriz de entrada compatível com o vetor de saída que, neste caso, contém os 664 valores da análise de laboratório. Desta forma, cada janela de entrada precisa ter suas dimensões reduzidas a um vetor [1x5], cujos pontos representarão os 360 instantes para cada um dos 5 instrumentos selecionados. Para isto, é calculada a média de alguns termos à esquerda e à direita do atraso máximo, determinado na correlação cruzada. Assim, o ruído sobre a entrada é reduzido e uma única matriz de entrada [664x5], obtida a partir da concatenação dos vetores encontrados para cada janela, é formada. Logo se tem uma matriz de entrada compatível com os padrões de saída,.

Para escolher a quantidade de termos para o cálculo da média, de forma a absorver a essência do sinal nos 360 instantes, não foi feita nenhuma heurística, mas sim, testes com 60, 120, 180, 240 e 300 termos. Isto gerou a necessidade de treinar as redes configuradas com cada uma destas cinco bases de dados, buscando a melhor resposta para o sistema.

Nesta fase também foram criados grupos de validação cruzada e de teste, com o objetivo de avaliar a eficiência da metodologia, conforme será visto no capítulo de resultados. Para isto, os dados de entrada e saída foram separados em grupos de treinamento, validação e teste, na proporção de 60%, 30% e 10%, respectivamente.

4.4. Inferência

Os sistemas de controle tradicional, como por exemplo, o controle proporcional integral e derivativo (PID), apesar de largamente aplicados na indústria, exigem uma alta complexidade na formação de modelos responsáveis pela interação das variáveis de monitoramento e controle instaladas na planta de

processo. Sendo assim, acabam sendo divididos em controles mais simples com malhas fechadas, que suprem a necessidade da operação como um todo.

Entretanto, um conceito relativamente novo vem conquistando seu espaço no mercado industrial dentro da área de controle avançado, a inferência de multivariável. Inferir uma informação é concluir algo a partir de dados que podem estar diretamente ou indiretamente relacionados a esta informação. Na prática, a técnica torna possível relacionar os dados de leitura, que monitoram a planta de um dado processo, e controlar múltiplas variáveis de saída, garantindo com eficiência a qualidade de um produto final.

Na metodologia proposta, após a obtenção dos padrões de treinamento, as habilidades de um aprendizado utilizando redes neurais artificiais foram combinadas com o uso do algoritmo do filtro de Kalman estendido, conforme ilustrado na Figura 19.

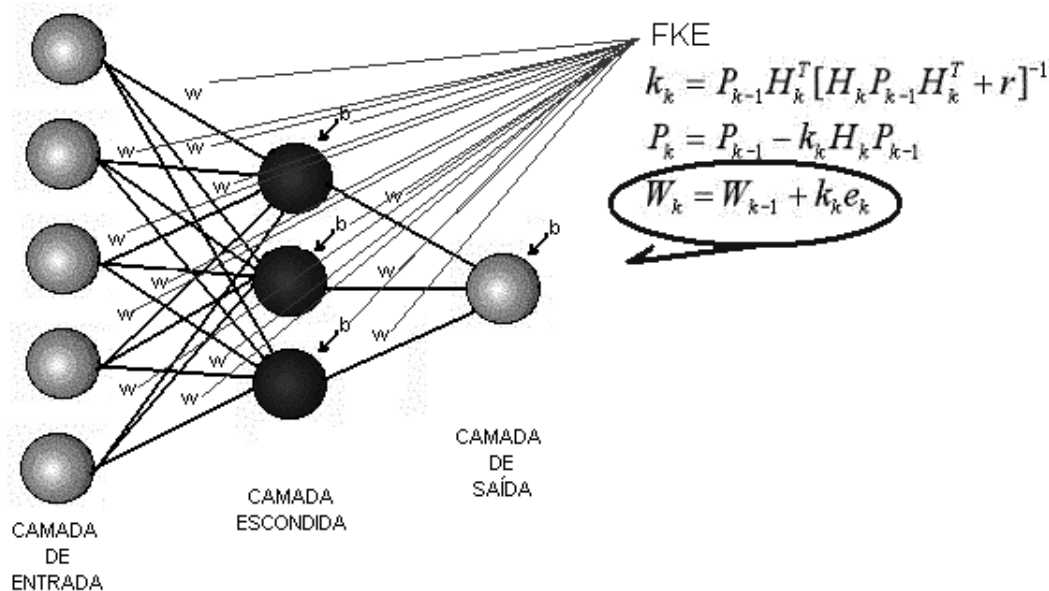


Figura 19: Estrutura da rede neural combinada com o FKE.

O sistema elaborado foi implementado em duas etapas independentes. Na primeira foram treinadas diversas redes neurais variando tanto o número de neurônios da camada escondida, quanto os grupos de dados de entrada e saída padrão, gerados na seção anterior. Na segunda, foi aplicado o algoritmo do filtro de Kalman estendido sobre os parâmetros ótimos apresentados por cada rede treinada. Desta forma, a atualização dos parâmetros da rede não foi executada por ela, mas sim, pelo filtro de Kalman estendido, que permite a adaptação do resultado da saída inferida sem necessidade de re-treinamento da rede.

O passo a passo do método combinando RNA e FKE está apresentado adiante, com base em uma das estruturas testadas no processo, de forma a esclarecer a implementação do sistema de inferência proposto.

Considere uma rede neural composta por cinco entradas de dados, que correspondem às cinco variáveis selecionadas no sub-item 4.2.3, e cinco neurônios na camada escondida, que representam as conexões ponderadas que extraem as características intrínsecas do processo de aprendizado. Esta estrutura se completa com um único neurônio de saída, que fornece as informações inferidas, que no nosso caso é a qualidade da nafta craqueada produzida.

1. Encontrada a rede neural ótima ajustada anteriormente com o grupo de dados de treinamento, capturam-se os parâmetros (pesos sinápticos e bias) da mesma armazenando-os em um vetor linha Ψ . Considerando o número de termos de entrada ($N_{in} = 5$), de saída ($N_{out} = 1$) e de neurônios da camada escondida ($N_{hid} = 5$), é possível definir a dimensão deste vetor calculando $N_{est} = N_{hid} \cdot (N_{in} + N_{out} + 1) + N_{out} = 36$.
2. São definidos também os parâmetros R e Q específicos do filtro de Kalman e relacionados ao tipo de amortecimento do sistema. Estes termos são constantes, porém, diversos valores foram testados para avaliar o comportamento do sistema.
3. Obtém-se uma matriz P_{k-1} de dimensão $[N_{est} \times N_{est}]$ com as estimativas iniciais de covariância através da relação $P_{k-1} = I \cdot \delta^1$, onde I é uma matriz identidade, δ é uma constante positiva e pequena e k é o instante da interação.
4. São apresentados dados do conjunto de teste à entrada da rede neural, gerando, assim, uma saída respectiva y_k , através dos pesos e bias armazenados na rede ótima inicial.
5. Com isso, calcula-se o erro e_k que representa a diferença entre a saída desejada d_k , armazenada no grupo de teste, e a saída y_k gerada pela rede ótima, isto é, $e_k = d_k - y_k$.

6. Calcula-se a matriz de medida H_k de dimensão $[N_{out} \times N_{est}] = [1 \times 36]$, que representa as derivadas parciais da saída em relação aos pesos e aos bias da rede. Para explicar a formação de H_k , é utilizada como exemplo a estrutura de rede apresentada na Figura 20, que apesar de não representar a rede em questão, envolve todos os cálculos necessários para a compreensão.

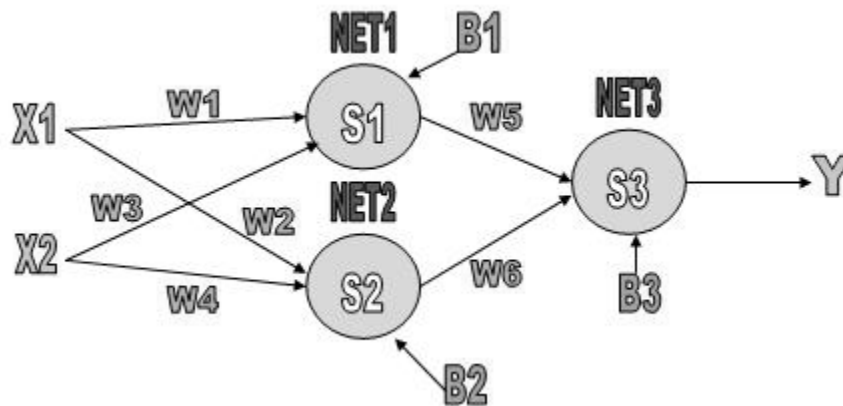


Figura 20: Exemplo de uma estrutura RNA.

- a. Para cada neurônio da rede neural é feita a combinação linear: $net_j = \sum x_i \cdot w_{ji} + b_j$. Desta forma, o resultado para o exemplo será a formação de:
- i. $net_1 = x_1 \cdot w_1 + x_2 \cdot w_3 + b_1$
 - ii. $net_2 = x_1 \cdot w_2 + x_2 \cdot w_4 + b_2$
 - iii. $net_3 = s_1 \cdot w_5 + s_2 \cdot w_6 + b_3$, onde s_1 e s_2 representam o resultado de net_1 e net_2 aplicadas à função de ativação f , ou seja, $s_1 = f(net_1)$ e $s_2 = f(net_2)$.
- b. Para o cálculo da saída da rede temos $y = s_3$, onde, $s_3 = f(net_3)$ ou ainda, $s_3 = f(s_1 \cdot w_5 + s_2 \cdot w_6 + b_3)$.
- c. Calculando a derivada parcial da saída y em relação a cada parâmetro da rede, é formada a matriz de medida H_k . Para este exemplo, N_{est} não será 36, mas sim, $N_{est} = 2 \cdot (2 + 1 + 1) + 1 = 9$, logo H_k terá dimensão $[1 \times 9]$ e cada um de seus elementos será dado por:
- i. $\partial y / \partial w_1 = \partial y / \partial s_1 \cdot s_1 / \partial w_1$
 - ii. $\partial y / \partial w_2 = \partial y / \partial s_1 \cdot s_1 / \partial w_2$
 - iii. $\partial y / \partial b_1 = \partial y / \partial s_1 \cdot s_1 / \partial b_1$

- iv. $\partial y / \partial w_3 = \partial y / s_2 * s_2 / \partial w_3$
 - v. $\partial y / \partial w_4 = \partial y / s_2 * s_2 / \partial w_4$
 - vi. $\partial y / \partial b_2 = \partial y / s_2 * s_2 / \partial b_2$
 - vii. $\partial y / \partial w_5 = \partial y / s_3 * s_3 / \partial w_5$
 - viii. $\partial y / \partial w_6 = \partial y / s_3 * s_3 / \partial w_6$
 - ix. $\partial y / \partial b_3 = \partial y / s_3 * s_3 / \partial b_3$
7. Por fim, o ganho de Kalman é calculado, através da Equação 8, e utilizado para encontrar a nova estimativa da matriz de correlação, conforme a Equação 9. Desta forma, os pesos sinápticos e bias da rede neural são atualizados pela Equação 10.

$$K_k = P_{k-1} \cdot H_k^T \cdot [H_k \cdot P_{k-1} \cdot H_k^T + R]^{-1} \quad (8)$$

$$P_k = P_{k-1} - K_k \cdot H_k \cdot P_{k-1} \quad (9)$$

$$\Psi_{P_k} = \Psi_{P_{k-1}} + K_k \cdot e_k \quad (10)$$

8. Para inferir uma nova saída a partir de um novo grupo de dados de entrada, o processo deve atualizar a matriz de covariância, substituindo P_{k-1} por P_k , carregar o vetor Ψ para a configuração da nova rede e substituir os valores a priori pelos valores a posteriori fazendo $\Psi_{k-1} = \Psi_k$. Desta forma, o sistema pode executar novamente o processo de inferência, retornando ao passo 4, sem necessidade de re-treinamento da rede neural, o que permite que todo o processamento seja em tempo real.